

An Experimental Study on E-commerce Customer Churn Prediction Based on the Explainable XGBoost Algorithm

Linye Li *

School of Business and Management, Jilin University, Changchun, Jilin, 130012, China

* Corresponding author Email: lily9923@mails.jlu.edu.cn

Abstract: The e-commerce industry has entered a phase of deepening market penetration, with high customer churn rates becoming a key issue constraining business development. Addressing the shortcomings of traditional churn prediction models—such as insufficient accuracy, poor interpretability, and failure to account for the heterogeneity of customer group needs—this paper proposes a framework for e-commerce customer churn prediction that integrates RFM value segmentation, the XGBoost algorithm, and SHAP interpretability analysis. Using a publicly available e-commerce dataset as the experimental sample, the RFM model was employed to segment customers into four value groups. Differentiated prediction models were constructed for each group, and SHAP was utilised to perform three-level interpretability analysis: global, group-level and local. Experimental results indicate significant heterogeneity in the churn drivers across the high-value, medium-value, potential-value and low-value customer sub-models. This study effectively addresses the ‘black box’ problem of models, offering enhanced interpretability and greater operational value, thereby providing support for e-commerce enterprises in making refined customer retention and operational decisions.

Keywords: Customer Churn Prediction; RFM Segmentation; XGBoost; SHAP Explainability; E-commerce Optimization.

1. Introduction

1.1. Research Background

With the rapid development of the digital economy, the e-commerce industry has transitioned from competing for incremental growth to a phase of deepening engagement with the existing customer base. According to industry statistics, the annual average user churn rate in the e-commerce sector has exceeded 30%, while the cost of re-acquiring churned customers is 5 to 10 times higher than that of acquiring new customers. Against a backdrop of continuously rising customer acquisition costs and increasingly fierce market competition, customer churn has become a core bottleneck constraining the profitability and sustainable development of e-commerce enterprises. Accurately identifying customers at high risk of churn, identifying the causes of churn, and formulating targeted retention strategies have become key challenges in customer relationship management for e-commerce enterprises.

Machine learning algorithms, with their powerful feature learning and predictive capabilities, have gradually replaced traditional statistical methods to become the mainstream tools for customer churn prediction. Among these, the XGBoost algorithm has been widely adopted in the field of e-commerce churn prediction due to its efficient gradient boosting mechanism, strong resistance to overfitting, and excellent adaptability to high-dimensional, imbalanced data [1]. However, traditional XGBoost models suffer from a significant ‘black box’ flaw: while they can output prediction results, they cannot explain the model’s decision-making logic or the drivers of churn, making it difficult for enterprises to formulate targeted retention strategies. Furthermore, most existing predictive models employ a ‘one-size-fits-all’ approach to modelling, overlooking the heterogeneity in consumption behaviour and preference among customers of

varying value. This results in insufficient predictive accuracy for the churn of core, high-value customers, making it difficult to meet enterprises’ needs for refined operational management.

1.2. Research Questions

Against this research background, this study focuses on addressing the following two core issues: first, how to achieve a deep integration of customer value segmentation with churn prediction models, thereby constructing personalised prediction models for different value segments to enhance prediction accuracy and relevance; second, how to optimise the XGBoost algorithm to achieve model interpretability, identify the key drivers of churn for different customer value segments, and provide clear evidence for the formulation of segmented retention strategies.

1.3. Current State of Research at Home and Abroad

Research in the field of customer churn prediction began earlier abroad, having developed a comprehensive research framework ranging from traditional statistical methods to ensemble learning and explainable artificial intelligence. The XGBoost algorithm, proposed by Chen and Guestrin, has rapidly become the mainstream algorithm for churn prediction in sectors such as e-commerce and telecommunications due to its excellent predictive performance [1]. In their research on telecommunications customer churn, Asif et al. combined XGBoost with the SHAP method to achieve both global and local interpretability analysis of the model, thereby validating the effectiveness of interpretable ensemble models. Regarding the integration of customer value segmentation and predictive modelling [2], Ansari and Schelle proposed a method combining RFM-KMeans segmentation with predictive

modelling, and validated the positive impact of segmentation strategies on improving predictive performance using online retail data [3].

Domestic scholars have focused on the localised optimisation of the XGBoost algorithm and its application in e-commerce scenarios, achieving a series of research outcomes with significant practical value. Wang Ying et al. conducted a systematic study on e-commerce customer churn prediction based on ensemble learning methods, verifying the effectiveness of ensemble learning in churn prediction scenarios [4]. Zuo Min et al. used Douyin e-commerce user data as their research subject and employed the Whale optimisation algorithm to optimise the core parameters of XGBoost; the churn prediction model they constructed significantly outperformed traditional machine learning models in terms of metrics such as AUC and F1 score [5]. Zhang Yuxuan proposed incorporating RFM segmentation results as features into the XGBoost predictive framework; through feature selection and the integration of segmentation features, the model's predictive accuracy was effectively enhanced [6].

Existing research has confirmed the superiority of the XGBoost algorithm in churn prediction and the application value of RFM segmentation in the e-commerce industry; however, significant shortcomings remain: firstly, most models lack deep integration between customer value stratification and predictive models, and a unified modelling approach struggles to accommodate the heterogeneous characteristics of different segments; secondly, research on interpretability largely remains at the level of global feature importance ranking, lacking segment-specific comparative explanations and precise individual-level explanations for segmented models; thirdly, there is a lack of research on the implementation of interpretable XGBoost models tailored to e-commerce scenarios, making it difficult to support enterprises' personalised operational requirements. This study aims to address these gaps.

1.4. Research Objectives

Centred on the prediction of e-commerce customer churn, this study constructs an integrated framework that combines RFM customer value segmentation, the XGBoost prediction algorithm, and the SHAP interpretability method. Specific objectives include: firstly, to achieve scientific value segmentation of e-commerce customers through the RFM model, thereby constructing personalised XGBoost churn prediction models for different segments to enhance prediction accuracy; secondly, to utilise the SHAP method to conduct three-level interpretability analysis (global, segment, and local) of the model, revealing the key drivers of churn for customers of different value; thirdly, to propose tiered customer retention strategies based on the research findings, providing feasible solutions for the refined operations of e-commerce enterprises.

2. Research Methodology

2.1. Theoretical Framework

RFM Customer Value Segmentation Theory: The RFM model uses Recency (time since last purchase), Frequency

$$R_score = \begin{cases} 4, & R \leq Q1 \\ 3, & Q1 < R \leq Q2 \\ 2, & Q2 < R \leq Q3 \\ 1, & R > Q3 \end{cases} \quad F_score = \begin{cases} 1, & F \leq Q1 \\ 2, & Q1 < F \leq Q2 \\ 3, & Q2 < F \leq Q3 \\ 4, & F > Q3 \end{cases} \quad (3)$$

(purchase frequency), and Monetary (spending amount) as core indicators. By quantifying customer activity, loyalty and value contribution, it enables customer value segmentation [7]. Due to its simplicity, ease of data acquisition and alignment with e-commerce operational realities, this model has become a classic tool in customer relationship management. This study employs the quartile method to determine thresholds for each indicator, categorising customers into four groups: high-value, medium-value, potential-value, and low-value.

Principles of the XGBoost Algorithm: The XGBoost algorithm is an optimisation algorithm based on gradient-boosted decision trees. It uses CART trees as base learners and achieves model optimisation by minimising the objective function through iterative training. Its objective function comprises a loss function and L1 and L2 regularisation terms, ensuring both predictive accuracy and effective suppression of overfitting. The algorithm supports features such as feature subsampling and sparse data optimisation, and demonstrates excellent adaptability to high-dimensional, imbalanced e-commerce data.

The XGBoost algorithm is an optimisation algorithm based on gradient-boosted decision trees. The objective function for the t th iteration is:

$$Obj^{(t)} = \sum_{i=1}^n L(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \sum_{k=1}^t \Omega(f_k) \quad (1)$$

SHAP explainability theory: The SHAP method is based on the Shapley value from game theory. The SHAP value of the j th feature in the i th sample is defined as:

$$\phi_j = \sum_{S \subseteq F \setminus \{j\}} \frac{|S|! (|F| - |S| - 1)!}{|F|!} [f_{S \cup \{j\}}(x_{S \cup \{j\}}) - f_S(x_S)] \quad (2)$$

By quantifying the marginal contribution of each feature to the prediction result, the method enables interpretability analysis of the model [8]. This method possesses a unified theoretical framework and is applicable to various machine learning models; it can both rank features by global importance and conduct group-based comparative explanations and individual prediction attribution.

2.2. Experimental Design

Data Preparation: The experimental data was sourced from the Kaggle E-commerce Customer Feature Dataset. Following merging and cleaning, a comprehensive dataset comprising 5,097 customer samples and 32 feature metrics was formed. The churn label was defined based on a combination of RFM scores and behavioural metrics. Data pre-processing included: handling outliers using the interquartile range method, encoding categorical features with a LabelEncoder, standardising continuous features with a MinMaxScaler, and finally, adaptively partitioning the training and test sets according to each cluster.

RFM Clustering Process: The core RFM metrics for each customer were calculated; the R, F, and M metrics were scored on a 1–4 scale using the quartile method (R was scored inversely, while F and M were scored positively); this study employed the quartile method to score the R, F, and M metrics on a 1–4 scale:

M_score and F_score follow the same scoring rules. The composite RFM score is calculated as:

$$RFM = R_score \times 100 + F_score \times 10 + M_score$$

$$RFM = R_score \times 100 + F_score \times 10 + M_score$$

Customers are categorised into four groups—high-value, medium-value, potential-value and low-value—based on their RFM scores.

Model Construction and Optimisation: Churn prediction models were constructed separately for each of the four RFM segments. For high-value, medium-value and potential-value customers, the XGBoost algorithm was employed, with hyperparameters optimised via 5-fold cross-validation; for low-value customers with a small sample size, logistic regression was used in conjunction with leave-one-out validation, and strong regularisation was applied to prevent overfitting.

Interpretability Analysis: The SHAP method was employed to conduct a three-level interpretability analysis:

identifying key churn features at the global level; comparing differences in churn drivers across different groups at the group level; and interpreting the predictive logic of typical individual samples at the local level. AUC was selected as the model evaluation metric, and the validity of the research methodology was verified by comparing the cluster-specific models with a unified model.

3. Research Results

3.1. Predictive Performance of Segmented Models

In this study, churn prediction models were constructed separately for four customer segments: high-value, Secondary value, potential-value and low-value. A validation strategy tailored to the sample size was adopted, Table 1 showed the performance of each segmented model.

Table 1. The performance of each segmented model

Segment	Sample Size	Churn Samples	Churn Rate	Validation Method	AUC
High value	680	223	32.79%	5-fold cross-validation	0.9663 (±0.0117)
Secondary value	997	236	23.67%	5-fold cross-validation	0.9781 (±0.0079)
Implied value	3318	1,235	37.22%	5-fold cross-validation	0.9562 (±0.0105)
Low value	102	81	79.41%	Leave-One-Out Method	0.9095
Unified Model	5097	1775	34.78%	5-fold cross-validation	0.9712 (±0.0085)

For customer segmentation across different sample sizes, differentiated model selection and parameter optimisation strategies were adopted. The focus was on optimising the XGBoost algorithm across four dimensions: regularisation, tree structure, sampling strategy and learning rate. This effectively prevented overfitting and enhanced the model’s generalisation ability. All customer segmentation models achieved excellent predictive performance. Specifically, the 5-fold cross-validation AUCs for the high-value customer model, medium-value customer model and potential-value customer model all remained stable within the range of 0.95 to 0.98, fully demonstrating that the segmentation modelling strategy possesses outstanding predictive performance and generalisation capabilities in scenarios with sufficient sample size; However, with only 102 low-value customers in the sample (including 21 non-churned customers), conventional cross-validation would result in an unstable training set distribution. Therefore, the leave-one-out method was employed for validation, yielding a model AUC of 0.9095. This demonstrates acceptable discriminatory power even under extreme sample constraints, thereby meeting the basic requirements for churn prediction.

For the three customer categories—high-value, medium-value and potential-value—we uniformly employed the XGBoost algorithm and configured targeted optimisation parameters. Specifically, for the regularisation dimension, we set `reg_alpha=1` and `reg_lambda=1`, introducing L1/L2 regularisation to control model complexity and prevent overfitting;For the tree structure parameters, we set `max_depth=2` and `n_estimators=50` to prevent the model from becoming overly complex by limiting the tree depth and

the number of iterations; for the sampling strategy parameters, we set `subsample=0.6` and `colsample_bytree=0.6` to enhance model diversity through random row and column sampling, thereby further improving generalisation ability;For the learning rate, `learning_rate` is set to 0.01; a low learning rate combined with an appropriate number of trees ensures more stable model convergence. Notably, after applying this optimised configuration to the high-value customer sub-model (sample size: 680 individuals), the AUC achieved via 5-fold cross-validation reached 0.9663 (±0.0117), validating the effectiveness of regularisation and sampling strategies in this scenario;The sub-model for secondary-value customers (sample size: 997) adopted the same configuration, achieving a 5-fold cross-validation AUC of 0.9781 (±0.0079), making it the cluster with the best predictive performance among the four groups, indicating that this parameter configuration can effectively capture the churn patterns of secondary-value customers;The high-potential customer sub-model (sample size: 3,318) also utilised this configuration, achieving an AUC of 0.9562 (±0.0105) in a 5-fold cross-validation, maintaining stable predictive capability even in the scenario with the largest sample size and highest churn rate.

For the sub-model targeting low-value customers (sample size: 102), where the sample size was too small to support XGBoost’s complex parameter configuration, we switched to a logistic regression algorithm and set strong regularisation parameters to accommodate the small-sample scenario. The specific parameters were `C=0.1` (regularisation strength), `penalty='l2'`, and `class_weight='balanced'` (to address class imbalance). Validated using the leave-one-out method, this

model achieved an AUC of 0.9095, demonstrating acceptable discriminatory power under extreme sample constraints. Overall, the three XGBoost models utilised a unified set of optimisation parameters; even with significant differences in sample size, the models' AUCs all remained above 0.95, validating the generality and robustness of this parameter combination in the context of e-commerce customer churn prediction. Meanwhile, the use of logistic regression for low-value customers due to sample size constraints fully demonstrates the necessity of matching model selection to sample size.

3.2. Results of SHAP Explanatory Analysis

The SHAP analysis reveals the heterogeneity among the four customer churn-driven cohorts and the differences in churn drivers. The global SHAP feature importance analysis shows that `engagement_score` is the most critical feature driving churn across all segments, with its average absolute SHAP value ranking highest among high-, secondary- and potential-value customers. The ranking of feature importance across different segments exhibits significant heterogeneity: for high-value customers, the ranking is `engagement_score` > `days_since_last_purchase` > `cart_abandonment_rate` > `browsing_frequency_per_week`, classifying them as 'experience-driven'; the ranking for secondary-value customers is broadly similar to that of high-value customers, but with engagement and purchase interval as the core factors, classifying them as 'engagement-driven'; the ranking for potential-value customers is `engagement_score` > `total_orders` > `days_since_last_purchase` > `browsing_frequency_per_week`, with engagement and purchase frequency as the core factors, classifying them as 'behavioural- and price-driven'; low-value customers are subject to sample size limitations; their core feature is `days_since_last_purchase`, and churn is driven almost exclusively by purchase interval, exhibiting the simple characteristic of 'inactivity leading to churn'. The SHAP values for high- and secondary-value customers fluctuate more significantly, indicating that these groups are more sensitive to changes in platform activity. This feature ranks as the second most important across high-, secondary- and potential-value customers, validating the conclusion that 'purchase activity is the core early warning signal for e-commerce customer churn'.

3.3. Comparison of Segmented and Unified Models

To validate the effectiveness of segmented modelling, this study compared the performance of the segmented model with that of a non-segmented unified model. To ensure a fair comparison, the unified model employed the same 5-fold cross-validation strategy as the segmented model. The AUC of the unified model's 5-fold cross-validation was 0.9712, while the weighted average AUC of the segmented model was 0.9625 (the weighted average of the CV AUCs for the high-, medium- and potential-value segments).

Although the unified model demonstrated a slight advantage in overall accuracy, the cluster-based model achieved differentiated attribution of churn drivers across cohorts through SHAP explainability analysis. This overcomes the limitation of the unified model—which 'can only predict but cannot explain'—and provides e-commerce enterprises with a comprehensive solution encompassing 'precise identification of churn risk, identification of the

causes of churn within specific cohorts, and the formulation of targeted retention strategies'. This represents the core practical value of cluster-based modelling compared to unified modelling.

4. Discussion

4.1. Analysis of the Validity of Segmented Modelling

The research findings confirm that a personalised modelling strategy based on RFM segmentation can effectively enhance the interpretability of churn prediction. The high-value customer sub-model utilised XGBoost with 5-fold cross-validation; the sub-model for medium-value customers demonstrated the best performance (AUC 0.9781), while the sub-model for potential-value customers maintained stable predictive capability even under conditions of the largest sample size and highest churn rate (AUC 0.9562). The core advantage of segmented modelling lies in its ability to accurately capture the differences in churn characteristics across different segments, thereby providing a basis for differentiated attribution in tiered operations.

4.2. Interpretation of Heterogeneity in Churn Drivers

The heterogeneity of customer churn drivers revealed by SHAP explainability analysis aligns closely with the operational logic of e-commerce customer value segmentation. High-value customers are in the mature consumption phase, exhibiting high sensitivity to platform service experience and shopping processes; their churn decisions tend to be 'experience-driven'. Secondary-value customers are in the growth phase of consumption, being more sensitive to changes in platform activity levels; their churn decisions tend to be 'activity-driven'. Potential-value customers are in the early stages of consumption, where both purchase frequency and average order value jointly influence their retention decisions; their churn decisions tend to be "behaviour and price-driven". Low-value customers exhibit simple consumption behaviour and offer the lowest return on retention investment; businesses may adopt low-cost retention strategies or appropriately abandon them.

Both customer segments utilise `engagement_score` and `days_since_last_purchase` as core churn drivers, validating the conclusion that "activity levels and purchase intervals are key early warning signals for e-commerce customer churn". This provides quantitative support for businesses to formulate tiered customer engagement strategies.

4.3. Theoretical and Practical Value of the Model

At the theoretical level, this study has constructed an integrated framework comprising "RFM clustering—XGBoost prediction—SHAP interpretation", thereby enriching the application system of interpretable machine learning in the field of marketing and filling a gap in research on the interpretability of clustering models. At the practical level, the research findings can be directly applied to the refined operations of e-commerce enterprises: by using the clustering model to accurately identify high-risk churn customers, utilising SHAP analysis to clarify the causes of churn, and formulating differentiated strategies for different customer segments.

4.4. Research Limitations and Directions for Improvement

This study has three limitations: firstly, the experimental data utilised a publicly available dataset, which may differ in feature distribution from real-world enterprise data, necessitating further validation of generalisation capabilities; secondly, the RFM model employed only the traditional three metrics, failing to incorporate multi-dimensional features such as platform interactions and user profiling; thirdly, the sample size of low-value customers was extremely small, and model stability requires improvement. Future research could be enhanced by collecting real-world enterprise data, expanding the dimensions of RFM metrics, and introducing intelligent optimisation algorithms.

5. Summary and Conclusion

5.1. Research Summary

This study focuses on the need for accuracy and interpretability in e-commerce customer churn prediction, conducting research on the integration of RFM clustering and interpretable XGBoost models. Through a series of experiments involving data pre-processing, RFM clustering, model construction and optimisation, and interpretability analysis, the effectiveness of the clustering modelling strategy and interpretability methods was validated. During the study, the performance differences between models for distinct RFM segments were systematically compared, revealing the heterogeneous characteristics of churn drivers across segments, thereby establishing a comprehensive research chain comprising ‘segmentation—prediction—interpretation—strategy’.

5.2. Key Conclusion

Firstly, RFM-based customer value segmentation provides an effective foundation for churn prediction. The personalised models constructed for different segments achieved CV AUCs of 0.9663, 0.9781 and 0.9562 for high-, medium- and potential-value customers respectively, while the AUC for the ‘retain one’ method for low-value customers reached 0.9095, validating the effectiveness of the segmentation modelling strategy.

Second, the SHAP method effectively addresses the “black box” limitations of XGBoost, enabling interpretability analysis at three levels: global, group, and local. The composite engagement score (engagement score) is the core driver of churn across all customer segments, while the number of days since the last purchase (days_since_last_purchase) is the second most important feature.

Thirdly, the churn drivers for different customer value segments exhibit significant heterogeneity: high-value customers are ‘experience-driven’, mid-value customers are ‘engagement-driven’, potential-value customers are ‘behaviour and price-driven’, while low-value customers display the simple characteristic of ‘inactivity leading to churn’, offering greater interpretability and higher operational value.

Fourthly, the combination of segmentation modelling and interpretability analysis provides e-commerce enterprises with a comprehensive solution encompassing “precise identification, root-cause diagnosis and targeted interventions”, thereby facilitating the implementation of

refined operational strategies.

5.3. Practical Recommendations

Differentiated customer retention strategies can be formulated for different RFM customer value segments. For high-value customers, an exclusive customer service system should be established to optimise after-sales response times and problem resolution rates; personalised benefit packages and member-exclusive activities should be introduced to enhance brand loyalty; and timely outreach and early warnings should be provided for low-activity, high-value customers. For medium-value customers, leverage big data analysis to push precise product recommendations and regularly distribute targeted coupons to increase purchase frequency and average order value. Simultaneously, optimise the platform’s interactive experience to boost customer engagement, and monitor shopping cart abandonment rates to trigger recovery mechanisms. For potential-value customers, price incentives such as new-customer-exclusive offers and group-buying discounts can be introduced to lower the barrier to entry. A tiered reward mechanism—‘Favourites → Add to Basket → Purchase’—should be designed to activate latent demand, while increasing purchase frequency to enhance platform stickiness. For low-value customers, low-cost retention methods such as coupon push notifications and SMS reminders should be adopted to avoid wasting resources; for customers at high risk of churn, retention efforts may be appropriately scaled back, with core resources focused on high-value and potential-value customers. Future research could further expand the dimensions of metrics, optimise model algorithms, and validate the effectiveness of real-world applications, thereby continuously refining the theoretical framework and practical solutions for e-commerce customer churn prediction and retention.

References

- [1] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). San Francisco, CA. <https://doi.org/10.1145/2939672.2939785>.
- [2] Asif, M., Khan, Z., & Ali, S. (2021). Telecom customer churn prediction using XGBoost and SHAP values. *Journal of King Saud University – Computer and Information Sciences*, 33, 768–778. <https://doi.org/10.1016/j.jksuci.2020.09.014>.
- [3] Ansari, A., & Schelle, H. (2019). Customer segmentation and churn prediction using RFM and K-means. *Journal of Retailing and Consumer Services*, 49, 345–353. <https://doi.org/10.1016/j.jretconser.2019.03.014>.
- [4] Wang, Y., & Li, J. (2020). A study on e-commerce customer churn prediction based on ensemble learning. *Computer Engineering and Applications*, 56, 221–227.
- [5] Zuo, M., & Zhang, Y. (2022). E-commerce customer churn prediction using XGBoost optimised by the whale optimisation algorithm. *Computer Engineering and Applications*, 58, 230–238.
- [6] Zhang, Y. X. (2021). An intelligent customer churn prediction method integrating RFM and XGBoost. *Journal of Computer Applications*, 38, 789–793.
- [7] Hughes, A. M. (1994). *Strategic database marketing*. McGraw-Hill.
- [8] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems 30* (pp. 4765–4774). Long Beach, CA. <https://doi.org/10.5555/3295222.3295309>.